(Structured codes in) Code based cryptology

Nicolas Sendrier

CRI Paris-Rocquencourt, équipe-projet SECRET

September 25, 2009

Academy contact forum, Brussels

Why code-based crypto ?

- Secure public-key cryptographic primitives
- Resistant to quantum computers

Features:

Good	Bad
• Fast and simple (often)	• Large keys
• Tight security reduction	• Unused (yet)

Today's point

Lately, several proposal have been made to reduce the key size

- AfricaCrypt 2009. Berger, Cayrel, Gaborit and Otmani
  Using quasi-cyclic alternant codes

- SAC 2009. Barreto and Misoczki
  Using dyadic Goppa codes

- We can easily measure the impact on key size.
- Can we measure the impact on security ?

Outline

Introduction

Security reduction

Practical attacks

$\longrightarrow$ how can we improve the systems?

Using structured codes

$\longrightarrow$ how does this affect security?

Conclusions

Introduction

Code-based one-way encryption in one slide

$\mathcal{C} \subset \{0,1\}^n$ a binary (linear) code $\qquad$ $\mathcal{E} \subset \{0,1\}^n$ a set of errors

$$\begin{array}{rcl} f: \ \mathcal{C} \times \mathcal{E} & \longrightarrow & \{0,1\}^n \\ (x, e) & \longmapsto & x + e \end{array}$$

$$\left.\begin{array}{l} \mathcal{C} \text{ has minimum Hamming distance } \geq 2t+1 \\ \mathcal{E} \text{ formed with words of Hamming weight } \leq t \end{array}\right\} \Rightarrow f \text{ is injective}$$

In general f is one-way (deciding $y \in f(\mathcal{C} \times \mathcal{E})$ is NP-complete)

Any (fast) t-bounded decoder for $\mathcal{C}$ provides a trapdoor

Code-based crypto - Main issues

- message security: decoding attacks
  $\longrightarrow$ decoding is hard in average (conjecture)
  $\longrightarrow$ finding a weakness is unlikely
  $\longrightarrow$ studying decoding attacks needed for maintenance

- key security: structural attacks
  $\longrightarrow$ which code family for which security ?
  $\longrightarrow$ can we harmlessly reduce the key size ?
  $\longrightarrow$ need for research

- what if we do not need a trap ?
  $\longrightarrow$ authentification, PRNG, hash function
  $\longrightarrow$ no structural attacks
  $\longrightarrow$ allows larger t

Syndrome mapping

$\mathcal{C}$ a binary linear (n, k) code
$H \in \{0,1\}^{r \times n}$ a parity check matrix of $\mathcal{C}$, $r = n - k$
$W_{n,t}$ the words of length n and Hamming weight t

Code-based cryptosystems rely on the "one-wayness" of the H-syndrome

$$\begin{array}{rccc} S_H : & W_{n,t} & \longrightarrow & \{0,1\}^r \\ & e & \longmapsto & eH^T \end{array}$$

Decoding in a binary linear code is equivalent to invert S_H, no more, no less

Syndrome decoding

$\mathcal{C}(n,k)$ a binary linear code
$H \in \{0,1\}^{r\times n}$ a parity check matrix, $r = n - k$

H-syndrome decoder

$$\begin{array}{rccc} \Psi_H: & \{0,1\}^r & \rightarrow & \{0,1\}^n \\ & s & \mapsto & e \end{array} \qquad \text{such that } s = eH^T$$

If $2t < \text{dmin}(\mathcal{C})$, Ψ_H is t-bounded if for all $e \in \{0,1\}^n$

$$\text{wt}(e) \le t \Rightarrow \Psi_H(eH^T) = e$$

More generally, Ψ_H is t-bounded if for all $e \in \{0,1\}^n$

$$\text{wt}(e) \le t \Rightarrow \text{wt}(\Psi_H(eH^T)) \le t$$

(if there are words of weight $\le t$ in a coset, the decoder finds one)

Syndrome decoding

$\mathcal{C}(n,k)$ a binary linear code
$H \in \{0,1\}^{r\times n}$ a parity check matrix, $r = n-k$

H-syndrome decoder

$$\begin{array}{rrcl} \Psi_H : & \{0,1\}^r & \rightarrow & \{0,1\}^n \\ & s & \mapsto & e \end{array} \qquad \text{such that } s = eH^T$$

If $2t < \mathrm{dmin}(\mathcal{C})$, Ψ_H is t-bounded if for all $e \in \{0,1\}^n$

$$\mathrm{wt}(e) \leq t \Rightarrow \Psi_H(eH^T) = e$$

More generally, Ψ_H is t-bounded if for all $e \in \{0,1\}^n$

$$\mathrm{wt}(e) \leq t \Rightarrow \mathrm{wt}(\Psi_H(eH^T)) \leq t$$

(if there are words of weight $\leq t$ in a coset, the decoder finds one)

Two instantiations of the code-based one-way function

n the code length	$\mathcal{C}(n,k)$ a binary linear code
k the dimension	$G \in \{0,1\}^{k \times n}$ a generator matrix
$r = n - k$ the codimension	$H \in \{0,1\}^{r \times n}$ a parity check matrix
t the error weight	$W_{n,t}$ the words of length n and weight t

encoding + noise	syndrome
$f_G : \{0,1\}^k \times W_{n,t} \rightarrow \{0,1\}^n$ $(x, e) \mapsto xG + e$	$S_H : W_{n,t} \rightarrow \{0,1\}^r$ $e \mapsto eH^T$

Both are equally hard to invert and can be inverted using a t-bounded (syndrome) decoder

Conversely, from f_G^{-1} or S_H^{-1}, we easily define a t-bounded decoder

An example: McEliece PKC (1978)

$\mathcal{C}$ a t-error correcting irreducible binary Goppa code of length 2^m

Parameters: $(m, t) \rightarrow$ length $n = 2^m$ and dimension $k = n - mt$

Public key: $G \in \{0, 1\}^{k \times n}$ a generator matrix of $\mathcal{C}$

Secret key: Ψ_H, a t-bounded H-syndrome decoder for any parity check matrix H of $\mathcal{C}$

Plaintext: $x \in \{0, 1\}^k$
Encryption: $x \mapsto xG + e$ with e a random error of weight t

Ciphertext: $y \in \{0, 1\}^n$
Decryption: $y \mapsto (y - \Psi_H(yH^T))G^*$ where $GG^* = 1 \in \{0, 1\}^{k \times k}$

Original parameters: $n = 1024$, $k = 524$ and $t = 50$

[McEliece, 1978]
"A public-key cryptosystem based on algebraic coding theory"

Another example: Niederreiter PKC (1986)

$\mathcal{C}$ is a t-error correcting binary linear (n, k) code

Parameters: length n, codimension $r = n - k$ and error weight t

Public key: $H \in \{0, 1\}^{r \times n}$ a parity check matrix of $\mathcal{C}$

Secret key: Ψ_H, a t-bounded H-syndrome decoder

Plaintext: $e \in W_{n,t}$
Encryption: $e \mapsto S_H(e) = eH^T$

Ciphertext: $s \in \{0, 1\}^r$
Decryption: $s \mapsto \Psi_H(s)$

[Niederreiter, 1986]
"Knapsack-type cryptosystems and algebraic coding theory"

Main code-based cryptosystem

Public key encryption: McEliece (1978); Niederreiter (1986)

Digital signature: Courtois, Finiasz, S. (2001)

PRNG: Fischer, Stern (1996)

Stream cipher: Gaborit, Laudaroux, S. (2007)

Hash function: FSB (2005); SHA3-FSB (2008)

Zero-knowledge: Stern (1993); Véron (1995); Gaborit, Girault (2007)

And also

- Rank metric (Gabidulin codes), weakened by Overbeck
- HB and its variants (low cost identification), also weakened
- . . .

Security reduction

Hard decoding problems

Syndrome Decoding NP-complete

Instance: $H \in \{0,1\}^{r \times n}$, $s \in \{0,1\}^r$, w integer

Question: Is there $e \in \{0,1\}^n$ such that $\text{wt}(e) \le w$ and $eH^T = s$?

Computational Syndrome Decoding NP-hard

Instance: $H \in \{0,1\}^{r \times n}$, $s \in \{0,1\}^r$, w integer

Output: $e \in \{0,1\}^n$ such that $\text{wt}(e) \le w$ and $eH^T = s$

Goppa Bounded Decoding NP-hard

Instance: $H \in \{0,1\}^{r \times n}$, $s \in \{0,1\}^r$

Output: $e \in \{0,1\}^n$ such that $\text{wt}(e) \le \dfrac{r}{\log_2 n}$ and $eH^T = s$

Open problem: average case complexity (Conjectured difficult)

Decoding adversary

For given parameters n, r and t

For any program $\mathcal{A} : \{0,1\}^r \times \{0,1\}^{r \times n} \to W_{n,t}$, we define the event

$$\mathcal{S}_{\mathcal{A}} = \{(e,H) \in \Omega \mid \mathcal{A}(eH^T, H)H^T = eH^T\}$$

in the sample space $\Omega = W_{n,t} \times \{0,1\}^{r \times n}$ uniformly distributed

$\mathcal{A}$ is a (T, ε)-decoder if

- running time: $|\mathcal{A}| \leq T$
- success probability: $\mathrm{Succ}(\mathcal{A}) = \mathrm{Pr}_{\Omega}(\mathcal{S}_{\mathcal{A}}) \geq \varepsilon$

Irreducible binary Goppa codes

Parameters: m, t and $n \leq 2^m$

Let $\begin{cases} L = (\alpha_1, \ldots, \alpha_n) \text{ distinct in } \mathbf{F}_{2^m} \\ g(z) \in \mathbf{F}_{2^m}[z] \text{ monic irreducible of degree } t \end{cases}$

The binary irreducible Goppa code $\Gamma(L, g)$ of *support* L and *generator* $g(z)$ is defined as the following subspace of $\{0, 1\}^n$

$$a = (a_1, \ldots, a_n) \in \Gamma(L, g) \Leftrightarrow R_a(z) = \sum_{j=1}^{n} \frac{a_i}{z - \alpha_j} = 0 \mod g(z)$$

- the dimension of $\Gamma(L, g)$ is $k \geq n - tm$
- the minimum distance of $\Gamma(L, g)$ is $d \geq 2t + 1$
- there exists a t-bounded polynomial time decoder for $\Gamma(L, g)$

Hard structural problems

Goppa code Distinguishing NP

Instance: $H \in \{0,1\}^{r \times n}$

Question: Is $\left\{x \in \{0,1\}^n \mid xH^T = 0\right\}$ a binary Goppa code?

Goppa code Reconstruction

Instance: $H \in \{0,1\}^{r \times n}$

Output: (L, g) such that $\Gamma(L, g) = \left\{x \in \{0,1\}^n \mid xH^T = 0\right\}$

- NP: the property is easy to check given (L, g)
- Completeness status is unknown
- Tightness: gap between decisional and computational problems

Goppa code distinguisher

For given parameters n, r

For any program $\mathcal{D} : \{0,1\}^{r \times n} \rightarrow \{\text{true}, \text{false}\}$, we define the events*

$$\begin{aligned} \mathcal{T}_{\mathcal{D}} &= \{H \in \Omega \mid \mathcal{D}(H) = \text{true}\} \\ \mathcal{G} &= \{H \in \Omega \mid H \in \mathcal{H}_{\text{goppa}}\} \end{aligned}$$

in the sample space $\Omega = \{0,1\}^{r \times n}$ uniformly distributed

$\mathcal{D}$ is a (T, ε)-distinguisher if

- running time: $|\mathcal{D}| \leq T$
- advantage: $\text{Adv}(\mathcal{D}) = \left| \text{Pr}_{\Omega}(\mathcal{T}_{\mathcal{D}}) - \text{Pr}_{\Omega}(\mathcal{T}_{\mathcal{D}} \mid \mathcal{G}) \right| \geq \varepsilon$

*$\mathcal{H}_{\text{goppa}}$ the set of all parity check matrices of a Goppa code

Adversary for McEliece

For given parameters n, r and t

For any program $\mathcal{A} : \{0,1\}^r \times \{0,1\}^{r\times n} \to W_{n,t}$, we define the events

$$\begin{aligned} \mathcal{S}_\mathcal{A} &= \{(e,H) \in \Omega \mid \mathcal{A}(eH^T, H)H^T = eH^T\} \\ \mathcal{G} &= \{(e,H) \in \Omega \mid H \in \mathcal{H}_{\text{goppa}}) \end{aligned}$$

in the sample space $\Omega = W_{n,t} \times \{0,1\}^{r\times n}$ uniformly distributed

$\mathcal{A}$ is a (T,ε)-adversary (for McEliece) if

- running time: $|\mathcal{A}| \leq T$
- success probability: $\text{Succ}_{\text{McE}}(\mathcal{A}) = \text{Pr}_\Omega(\mathcal{S}_\mathcal{A} \mid \mathcal{G}) \geq \varepsilon$

If there exists a (T,ε)-adversary then there exists either
- a $(T,\varepsilon/2)$-decoder,
- or a $(T + O(n^2), \varepsilon/2)$-distinguisher,

Security reduction

Assuming

- decoding in a random linear code is hard
- Goppa codes are pseudorandom

McEliece cryptosystem is a One Way Encryption (OWE) scheme.

Using the proper semantically secure conversion any deterministic OWE scheme can become IND-CCA2

[Biswas, S. 2008] Without loss of security:

- McEliece's scheme can be made deterministic (by encoding information in the error)
- the public key can be in systematic form

[Kobara, Imai 2001] First IND-CCA2 conversion for McEliece

Practical security

Best known attacks

Decoding attacks: variants of information set decoding [Stern 1989]

Stern 1989; Canteaut, Chabaud 1998; Bernstein, Lange, Peters 2008

bounds: Bernstein, Lange, Peters, van Tilborg 2009; Finiasz, S. 2009

also (for large t): Wagner's Generalized Birthday Attack (2002)

Structural attacks: support splitting algorithm [S. 2000]

$\longrightarrow$ find the permutation between equivalent codes in polynomial time

McEliece/Niederreiter cryptosystem - Parameters

Using binary irreducible Goppa codes

(m,t)	sizes					security (in bits)	
	McEliece		Niederreiter		public key (syst.)		
	block	info	block	info		dec.	struct.
$(10,50)$	1024	524	500	284	32 kB	60	491
$(11,32)$	2048	1696	352	233	73 kB	86	344
$(12,40)$	4096	3616	480	320	212 kB	127	471

Can we trade some of the extra key security for a smaller key size?

Which family of codes for McEliece/Niederreiter systems

Should not be used

- Generalized Reed-Solomon codes (Sidelnikov, Shestakov 1992)
- Concatenated codes (S. 1998)
- Reed-Muller codes (Minder, Shokrollahi 2007)
- Algebraic geometry codes of low genus (Faure, Minder 2008)
- Turbo-codes, LDPC codes

Unbroken so far

- Goppa codes

New trend: structured codes (Gaborit 2005)

- Allow smaller key size
- Security reduction has to be revised

Structured codes

Using structured codes without trapdoor

Idea: the parity check matrix H is randomly chosen circulant by block. The whole matrix is defined by only a single or a few rows.

For such matrices, syndrome decoding remains NP-complete.

(Well chosen) quasi-cyclic codes meet the Gilbert-Varshamov bound.

$\rightarrow$ It is likely that PRNG, hash functions or zero-knowledge scheme will be as secure with random quasi-cyclic codes as with random codes.

Used in:

- Gaborit and Girault zero-knowledge protocol (2007)
- SYND stream cipher (2007)
- SHA3-FSB hash function (2008)

Structured codes for PKC

Idea: the secret code is cyclic or quasi-cyclic and the code positions are shuffled using a structured permutation. The resulting public key is structured and is defined by only a single or a few rows.

Security reduction now requires:

- decoding in a random quasi-cyclic code is hard (NP-complete)
- the public code is indistinguishable from a random quasi-cyclic code

The story

- First proposition with quasi-cyclic codes by Gaborit in 2005
- Broken by Otmani and Tillich in 2008
- Second quasi-cyclic proposal by Berger, Cayrel, Gaborit and Otmani in 2009
- Broken by Faugère, Otmani and Perret, last week
- Another similar idea using dyadic Goppa codes by Barreto and Misoczki in 2009
- ...

Conclusions

- Random structured codes are probably an excellent alternative to random codes
- Structured codes for PKC are another matter
- Anything else than binary Goppa codes seems to have flaws
- We need more research on structural attacks
  $\longrightarrow$ new families of codes
  $\longrightarrow$ new key reduction techniques

Can we trade some of the extra key security for a smaller key size?

I don't know!

Thank you